Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Hydrology 297 (2004) 236–255
Journal of Hydrology
www.elsevier.com/locate/jhydrol

A spatial time series framework for simulating daily precipitation at regional scales

P.C. Kyriakidis[a,*], N.L. Miller[b], J. Kim[c]

[a]*Department of Geography, University of California, Ellison Hall 5710, Santa Barbara, CA 93106, USA*
[b]*Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA*
[c]*Department of Atmospheric Sciences, University of California, Los Angeles, CA, USA*

Received 27 August 2002; revised 5 April 2004; accepted 16 April 2004

Abstract

A spatiotemporal stochastic simulation approach for constructing maps of daily precipitation at regional scales in a hindcast mode is proposed in this paper. Parametric temporal trend models of precipitation are first established at the available rain gauges. Temporal trend model parameters are then regionalized in space accounting for their spatial auto- and cross-correlation, as well as for their relationships with auxiliary spatial information such as terrain elevation. The resulting residual values at the rain gauges are modeled as a realization of a stationary spatiotemporal process. Sequential simulation is then used to generate alternative synthetic realizations of daily precipitation fields, which reproduce: (i) the rain gauge measurements, and (ii) their histogram and a model for their spatiotemporal correlation over the entire study region and time period of interest. In addition, a post-processing transformation allows reproduction of the rainfall histogram at particular dates, while preserving the observed rain gauge data.

A case study illustrates the applicability of the proposed methodology using daily precipitation measurements recorded at 77 rain gauges in the northern California coastal region from Nov 1, 1981 to Jan 31, 1982. Conditional stochastic simulation in space and time is performed for generating a 30-member ensemble of daily precipitation fields on a 300×360 grid of cell size 1 km^2 for the above time period. It is shown that simulated precipitation fields reproduce the spatiotemporal characteristics of the rain gauge measurements, thus providing realistic inputs of precipitation forcing for hydrologic impact assessment studies.

Keywords: Rainfall modeling; Geostatistics; Cokriging; Spatial time series; Conditional stochastic simulation; Daily precipitation; Digital elevation model; NCAR/NCEP reanalysis

1. Introduction

Regional scale precipitation predictions constitute one of the most important input parameters for hydrologic impact assessment studies; the improvement of such predictions poses a very important research theme in hydrology (Entekhabi et al., 1999). A means for enhancing the accuracy of regional scale precipitation predictions is provided by limited area models (LAMs) (Giorgi and Mearns, 1991; Kim and Soong, 1996; Miller and Kim, 1996;

* Corresponding author.
E-mail address: phaedon@geog.ucsb.edu (P.C. Kyriakidis).

0022-1694/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2004.04.022

Kim et al., 1998). Dynamical downscaling using LAMs yields precipitation predictions, which are physically and dynamically consistent with other atmospheric variables produced in the downscaling procedure. Dynamical downscaling, however, is computationally expensive and not error-free due to limited spatial resolution and model parameterizations. Stochastic characterization of rainfall fields based on rain gauge data and ancillary information, e.g. terrain elevation, still provides one of the basic tools for constructing rainfall maps at regional scales (Bras and Rodríguez-Iturbe, 1985; Seo et al., 2000; Kyriakidis et al., 2001b), even though physical and dynamic consistency among downscaled variables is not guaranteed.

Time domain approaches for modeling daily precipitation are often based on time series models, e.g. multivariate autoregressive (AR) models (Bras and Rodríguez-Iturbe, 1985), or generalized linear models (Chandler and Wheater, 2002). Such approaches exploit the typically better informed time domain, but are limited to predictions only at rain gauge locations (Wilks, 1998; von Storch and Zwiers, 1999). This limitation hinders the all-important task of spatiotemporal mapping. Recently, time series approaches (in the form of weather generators) have been generalized to a continuous spatial domain and maps of precipitation levels are constructed at any arbitrary location via interpolation of time series model parameters (Hutchinson, 1995; Johnson et al., 2000).

In this paper, a framework for stochastic spatiotemporal simulation of daily precipitation in a hindcast mode is developed following Kyriakidis and Journel (2001). Observed precipitation levels in space and time are modeled as a joint realization of a collection of space-indexed time series, one for each spatial location. Time series model parameters are spatially varying, thus capturing space–time interactions. The residuals from these local trend models are regarded as a realization of a stationary spatiotemporal process. Stochastic simulation, i.e. the procedure of generating alternative precipitation realizations (synthetic fields) over the space–time domain of interest (Deutsch and Journel, 1998), is employed for ensemble prediction. Realizations of the residual process are generated via conditional stochastic simulation and added to the estimated trend component to produce alternative conditional realizations of the spatiotemporal distribution of daily precipitation. The simulated daily precipitation fields reproduce a data-based histogram and spatiotemporal covariance model, and identify the measured precipitation values at the rain gauges (conditional simulation); this latter characteristic distinguishes this approach from weather generators which are essentially unconditional stochastic simulations. The resulting synthetic precipitation fields can be used in a Monte Carlo framework for risk analysis studies in hydrologic impact assessment investigations (Bras and Rodríguez-Iturbe, 1985; Kyriakidis et al., 2001a).

The proposed methodology is developed in Section 2 in the following sequence: construction of parametric temporal trend models at all rain gauges (Section 2.1), regionalization of resulting parameters in space (Section 2.2), and simulation of the spatiotemporal precipitation field (Section 2.3). In Section 3, a case study is presented using a daily precipitation data set near the northern California coastal region from Nov 1, 1981 to Jan 31, 1982. Last, a brief discussion regarding potential improvements and future work is given in Section 4.

2. Spatial time series

Let $\{Z_{\mathbf{u}}(t),\ t \in T\}$ denote a space-indexed random process (a time series model), where T denotes the time span of interest, and $\mathbf{u} = (u_1, u_2)$ denotes the 2D coordinate vector (e.g. longitude and latitude) of an arbitrary location in the study domain D. The set of all possible precipitation profiles in the study region for the particular study period is modeled as a collection of spatially correlated time series models, one for each location $\mathbf{u}$:

$$\{Z_{\mathbf{u}}(t),\ t \in T\},\ \mathbf{u} \in D \tag{1}$$

where the subscript $\mathbf{u}$ emphasizes the spatial time series viewpoint adopted in this work.

The random variable $Z_{\mathbf{u}}(t)$ modeling the uncertainty about the unknown precipitation value $Z_{\mathbf{u}}(t)$ at the t-th time instant at a location $\mathbf{u}$ is decomposed into:

$$Z_{\mathbf{u}}(t) = m_{\mathbf{u}}(t) + R_{\mathbf{u}}(t),\ t \in T \tag{2}$$

where $m_{\mathbf{u}}(t)$ is a deterministic space–time component characterizing some 'average' smooth variability of

the spatiotemporal process, and $R_{\mathbf{u}}(t)$ is stationary, zero mean, residual component modeling higher frequency fluctuations around that trend in both space and time. Notation-wise, $E\{Z_{\mathbf{u}}(t)\} = m_{\mathbf{u}}(t)$ and $E\{R_{\mathbf{u}}(t)\} = 0$, i.e. the trend component is the expected value of the process, and $\mathrm{Cov}\{m_{\mathbf{u}}(t),\ R_{\mathbf{u}}(t)\} = 0$, i.e. the trend and residual components are independent since the former is considered deterministic.

The trend component typically characterizes long-term temporal patterns, for example precipitation variability attributed to regional climatic factors. Other patterns of variability, e.g. those linked to local weather conditions, are typically accounted for by the stochastic residual component. It should be stressed that the dichotomy of Eq. (2) is a (subjective) modeling decision: there is no 'true' temporal trend component, since there are no trend data. Two researchers, for example, might disagree on the particular climatic factors that give rise to long-term temporal patterns and should thus be linked to the trend component. The resulting residual component is therefore a collective term for all components of variability that are not included in the trend model (Thiébaux, 1997).

At regional scales, the temporal characteristics of precipitation profiles are not stationary in space; this entails, for example, that the corresponding temporal trend components $\{m_{\mathbf{u}}(t),\ t \in T\}$ and $\{m_{\mathbf{u}'}(t),\ t \in T\}$ at two different locations $\mathbf{u}$ and $\mathbf{u}'$ will differ according to the climatic conditions prevailing at these locations. Precipitation characteristics at a location $\mathbf{u}$ near, say, the ocean are expected to be different from those at a location $\mathbf{u}'$ in an orographically isolated area. It is therefore critical to consider spatially non-stationary patterns of temporal variability in the modeling procedure, as well as to account for the influence of ancillary information on the spatial distribution of these characteristics.

To this end, local parametric models for the temporal trend of daily precipitation are first established at the available rain gauges. The resulting temporal trend model parameters are then regionalized (interpolated) in space, accounting for: (i) their relationships with auxiliary spatial variables, such as terrain elevation and its interaction with large-scale specific humidity derived from an assimilated data product from the National Centers for Environmental Prediction and the National Center for Atmospheric Research (NCEP/NCAR reanalysis), see Kalnay et al. (1996), and (ii) their spatial auto- and cross-correlation via cokriging (Wackernagel, 1995).

2.1. Station-specific temporal trend models

Let $\mathbf{z}_\alpha = [z_\alpha(t_i),\ i = 1,\ldots,T_\alpha]'$ denote the $(T_\alpha \times 1)$ vector of precipitation measurements at the α-th rain gauge with coordinate vector $\mathbf{u}_\alpha$; here T_α denotes the number of available measurements (length of precipitation profile) at location $\mathbf{u}_\alpha$, and superscript $'$ denotes transposition. The sample profile $\mathbf{z}_\alpha$ at the α-th monitoring station is regarded as a realization of a random process $\{Z_\alpha(t_i),\ t_i \in T\}$, and is decomposed as:

$$\mathbf{z}_\alpha = \mathbf{m}_\alpha + \mathbf{r}_\alpha \tag{3}$$

where $\mathbf{m}_\alpha = [m_\alpha(t_i),\ i = 1,\ldots,T_\alpha]'$ denotes a $(T_\alpha \times 1)$ vector comprising the temporal trend at location $\mathbf{u}_\alpha$, and $\mathbf{r}_\alpha = [r_\alpha(t_i),\ i = 1,\ldots,T_\alpha]'$ denotes a $(T_\alpha \times 1)$ random vector comprising a realization of a zero mean, stochastic residual component at the same location.

The temporal trend component $\mathbf{m}_\alpha$ at the α-th rain gauge is modeled as a weighted linear combination of $(K+1)$ elementary profiles (signals):

$$\mathbf{m}_\alpha = \mathbf{F}\mathbf{b}_\alpha \text{ or equivalently}$$

$$m_\alpha(t_i) = \sum_{k=0}^{K} b_k(\mathbf{u}_\alpha) f_k(t_i),\ i = 1,\ldots,T_\alpha \tag{4}$$

where $\mathbf{F} = [\mathbf{f}_k,\ k = 1,\ldots,K]$ is a $(T_\alpha \times (K+1))$ temporal design matrix whose k-th column comprises the k-th elementary temporal profile $\mathbf{f}_k = [f_k(t_i),\ i = 1,\ldots,T_\alpha]'$, and $\mathbf{b}_\alpha = [b_k(\mathbf{u}_\alpha),\ k = 0,\ldots,K]'$ is a $(K+1)$ column vector of coefficients (intensities) associated with these profiles. Note that, by convention,

$$\mathbf{f}_0 = [\underbrace{1,\ldots,1}_{T_\alpha}]',$$

i.e. the 0-th elementary profile is the $T_\alpha \times 1$ unit vector.

Each elementary temporal profile $\mathbf{f}_k$ contributing to the trend component, is independent of the spatial location $\mathbf{u}_\alpha$ and should ideally have a physical interpretation pertinent to the entire study region. Such profiles could be parametric functions of time linked to stationary (in space) climatic signals at

regional scales. Periodicities, for example, could be handled via sinusoidal forms of elementary profiles (thus leading to a Fourier series decomposition). Alternatively, these elementary profiles could be identified to a set of orthogonal factors derived via empirical orthogonal function (EOF) analysis of the rain gauge precipitation profiles (Rao and Hsieh, 1991; Hisdal and Tveito, 1992; Sauquet et al., 2000). A simpler approach, which is adopted in the case study of this paper, amounts to considering only two elementary profiles: $\mathbf{f}_0$ and $\mathbf{f}_1$, with $\mathbf{f}_1$ taken as the spatial average of all sample precipitation profiles available in the study region. In all the cases listed above, the temporal trend component is a consequence of subjective decisions regarding the number, and shape (parametric or not) of the elementary profiles comprising the columns of matrix $\mathbf{F}$ in Eq. (4).

It should be noted here that, when the elementary profiles are derived solely from sample data, the resulting trend and residual components can only be used in a hindcast mode. Forecasting requires that these elementary profiles be first projected in the future, or that they be linked to other variables for which forecasts can be readily obtained. In these cases, one would also be looking at time-varying temporal trend coefficients, which should also be projected in the future. Such a forecast modeling approach is not addressed in this work.

Irrespective of their definition, these common regional signals are modulated locally, via the coefficient vector $\mathbf{b}_\alpha$, to define temporal trend components that are different from one rain gauge to another. This vector $\mathbf{b}_\alpha$ of temporal trend coefficients can be determined at each rain gauge location $\mathbf{u}_\alpha$, independently from one location to another, using multiple (linear or non-linear) regression. In this work, the vector $\mathbf{b}_\alpha$ at any rain gauge location $\mathbf{u}_\alpha$ is computed via ordinary least squares (OLS) as (Searle, 1971):

$$\mathbf{b}_\alpha = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{z}_\alpha \tag{5}$$

Once the rain gauge specific vector $\mathbf{b}_\alpha$ of temporal trend coefficients is computed at each rain gauge location $\mathbf{u}_\alpha$, the temporal trend $\{m_\alpha(t_i),\ t_i \in T_\alpha\}$ at that location is given by Eq. (4), and the corresponding residual profile is computed as:

$$\mathbf{r}_\alpha = \mathbf{z}_\alpha - \mathbf{F}\mathbf{b}_\alpha, \text{ or equivalently}$$

$$r_\alpha(t_i) = z_\alpha(t_i) - \sum_{k=0}^{K} b_k(\mathbf{u}_\alpha) f_k(t_i),\ i = 1, \ldots, T_\alpha \tag{6}$$

The temporal trend b-coefficients are essentially defined via the algorithm adopted for their computation, and they are subsequently treated as precise data. These rain-gauge derived coefficients are used to estimate (interpolate) the vector $\mathbf{b}_\mathbf{u}$ of unknown coefficients at any ungauged location $\mathbf{u}$, accounting for relevant ancillary spatial information such as terrain elevation (Sauquet et al., 2000; Bierkens et al., 2001). The term deterministic is used for these temporal trend coefficients, because their uncertainty at the rain gauge locations, as well as their uncertainty at ungauged locations due to their spatial interpolation, is not accounted for in this work. In what follows, we describe a method to account for the interaction (correlation) between these coefficients during their spatial interpolation.

2.2. Regionalizing temporal trend coefficients

Even if temporal trend models are established independently at each rain gauge, the resulting temporal trend b-coefficients are auto- and cross-correlated (covary) in space since they are derived from the same process z-data, themselves correlated in space and time. Such b-coefficients are not pair-wise cross-correlated at the zero spatial lag, if one uses EOF analysis of the n sample precipitation profiles to arrive at the temporal trend components (Rao and Hsieh, 1991). The spatial orthogonality of such EOF-derived coefficients, however, is not guaranteed for non-zero spatial lags (Goovaerts, 1993). Consequently, a more general approach consists of accounting for spatiotemporal interactions between the $(K+1)$ temporal trend components via the spatial (cross) correlation of the local trend b-coefficients, as well as for any relationship with pertinent spatial variables, such as terrain elevation and/or specific humidity.

The proposed procedure of regionalizing (interpolating in space) the temporal trend b-coefficients

proceeds in the following steps:

1. Spatial regression of the temporal trend b-coefficients derived at the n rain gauges on collocated values of auxiliary variables, such as terrain elevation or lower-atmosphere variables derived, say, from NCEP/NCAR reanalysis. This step yields a set of $(K+1)$ spatial regression residual values at any rain gauge location $\mathbf{u}_\alpha$, as well as a set of $(K+1)$ estimated b-coefficients at any ungauged location $\mathbf{u}$. In this work, samples of the auxiliary variables are assumed representative of an area equal to the cell size of the prediction/simulation grid, and their support (scale) differences with the rain gauge data are not accounted for.
2. Cokriging of the rain gauge residuals from the above spatial regression to compute a set of $(K+1)$ predicted residuals at any ungauged location $\mathbf{u}$. This step accounts for the spatial auto- and cross-correlation of the spatial regression residuals obtained at the rain gauges.
3. Reconstruction of the $(K+1)$ sets of b-coefficients at any ungauged location $\mathbf{u}$, by adding the estimated b-coefficients obtained from spatial regression (step 1), and the predicted residuals obtained from cokriging (step 2). This step leads to an estimated trend component at any ungauged location $\mathbf{u}$ via a modified version of Eq. (4), see hereafter.

More precisely, the n values of the k-th coefficient obtained at the n rain gauge locations are expressed as:

$$\mathbf{b}^{(k)} = \mathbf{G}_n \mathbf{d}_k + \mathbf{e}_k \tag{7}$$

where $\mathbf{b}^{(k)} = [b_k(\mathbf{u}_\alpha),\ \alpha = 1,\ldots,n]'$ is a $(n \times 1)$ column vector containing the values of the k-th coefficient at the n rain gauges, $\mathbf{G}_n = [\mathbf{g}_l,\ l = 0,\ldots,L]$ is a $(n \times (L+1))$ spatial design matrix whose l-th column contains n values of the l-th auxiliary spatial variable (e.g. terrain elevation) $\mathbf{g}_l = [g_l(\mathbf{u}_\alpha),\ \alpha = 1,\ldots,n]'$ at the n rain gauges, $\mathbf{d}_k = [d_k(\mathbf{u}_\alpha),\ k = 0,\ldots,K]'$ is a $(L+1)$ column vector of spatial regression coefficients, and $\mathbf{e}_k = [r_k(\mathbf{u}_\alpha),\ \alpha = 1,\ldots,n]'$ is a $(n \times 1)$ column vector of spatial regression residuals. Note that $\mathbf{b}^{(k)}$ should not be confused with $\mathbf{b}_\alpha$, the latter denoting the $(K+1)$ column vector of temporal trend coefficients at the α-th rain gauge.

In this work, the vector $[\mathbf{b}^{(k)}]^*$ of spatial regression predictions for the k-th temporal trend coefficient is computed via OLS as (Searle, 1971):

$$[\mathbf{b}^{(k)}]^* = \mathbf{G}_n \mathbf{d}_k = \mathbf{G}_n[(\mathbf{G}_n'\mathbf{G}_n)^{-1}\mathbf{G}_n'\mathbf{b}^{(k)}] \tag{8}$$

The resulting $(n \times 1)$ vector $\mathbf{e}_k$ of spatial regression residuals (at the n rain gauges) for the k-th temporal trend coefficient is then computed as:

$$\mathbf{e}_k = [\mathbf{b}^{(k)}]^* - \mathbf{G}_n \mathbf{d}_k \tag{9}$$

these residuals have, by construction, a zero mean.

If prediction is performed at P prediction locations, the $(P \times 1)$ vector $[\mathbf{b}_P^{(k)}]^*$ of spatial regression predictions for the k-th temporal trend coefficient is computed as:

$$[\mathbf{b}_P^{(k)}]^* = \mathbf{G}_P \mathbf{d}_k = \mathbf{G}_P[(\mathbf{G}_n'\mathbf{G}_n)^{-1}\mathbf{G}_n'\mathbf{b}^{(k)}] \tag{10}$$

where $\mathbf{G}_P$ is a $(P \times (L+1))$ spatial design matrix whose l-th column contains P values of the l-th auxiliary spatial variable (e.g. terrain elevation) $\mathbf{g}_l = [g_l(\mathbf{u}_p),\ p = 1,\ldots,P]'$ at the P prediction locations. The predicted value of the k-th unknown temporal trend coefficient $b_k(\mathbf{u})$ at an arbitrary ungauged location $\mathbf{u}$ derived via the above spatial regression procedure is denoted as $b_k^*(\mathbf{u})$.

In the general case, the spatial regression residuals computed from Eq. (8) exhibit spatial auto- and cross-correlation. Consequently, their spatial prediction calls for inferring the cross-covariance matrix of the vector random function (RF) $\{E_k(\mathbf{u}),\ \mathbf{u} \in D\},\ k = 0,\ldots,K$, modeling the joint spatial correlation of these residuals. The geostatistical algorithm of cokriging is adopted for this joint prediction task (Wackernagel, 1995). It should be stressed here that (co)kriging can be viewed as a deterministic interpolator, given the realization of the sample data, and has close connections with interpolation using splines and radial basis functions (Wackernagel, 1995). Consequently, the interpolated temporal trend b-coefficients are hereafter treated as deterministic, thus entailing that the resulting temporal trend component is also deterministic. The case of random temporal trend b-coefficients, although tractable, is not addressed in this paper.

The simple cokriging (SCK) estimate $e_0^*(\mathbf{u})$ for the unknown regression residual for the b_0-coefficient $e_0(\mathbf{u}) = b_0(\mathbf{u}) - b_0^*(\mathbf{u})$, for example, at any location $\mathbf{u} \in D$ is expressed as:

$$e_0^*(\mathbf{u}) = \sum_{k=0}^{K} \mathbf{w}_{0k}' \mathbf{e}_k \tag{11}$$

where $\mathbf{w}_{0k} = [w_{0k}(\mathbf{u}_\alpha),\ \alpha = 1, \ldots, n]'$ denotes the $n \times 1$ vector of cokriging weights assigned to the known regression residuals of vector $\mathbf{e}_k$ for prediction of the unknown regression residual $e_0(\mathbf{u})$ at location $\mathbf{u}$, and obtained per solution of the SCK system of equations:

$$\begin{bmatrix} \mathbf{C}_{00} & \cdots & \mathbf{C}_{0K} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{K0} & \cdots & \mathbf{C}_{KK} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_{00} \\ \vdots \\ \boldsymbol{w}_{0K} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{00} \\ \vdots \\ \mathbf{c}_{0K} \end{bmatrix} \tag{12}$$

where $\mathbf{C}_{k'k}$ denotes the $n \times n$ matrix of auto- or cross-covariance values $C_{E_{k'}E_k}(\mathbf{u}_\beta - \mathbf{u}_\alpha)$ between any pair of regression residuals $e_{k'}(\mathbf{u}_\beta)$ and $e_k(\mathbf{u}_\alpha)$, and $\mathbf{c}_{0k}$ denotes the $(n \times 1)$ column vector of auto- or cross-covariance values $C_{E_0E_k}(\mathbf{u} - \mathbf{u}_\alpha)$ between any unknown regression residual $e_0(\mathbf{u})$ and any residual value $e_k(\mathbf{u}_\alpha)$. Similar equations can be written for the spatial prediction of residuals related to other b_k-coefficients, i.e. for $k \neq 0$, see Goovaerts (1997) for further details.

The final combined estimate of the unknown k-th temporal trend coefficient $b_k(\mathbf{u})$ at any location $\mathbf{u} \in D$ is obtained as:

$$b_k^{**}(\mathbf{u}) = b_k^*(\mathbf{u}) + e_k^*(\mathbf{u}) \tag{13}$$

and it accounts for: (i) the relationship of the temporal trend coefficient with auxiliary spatial variables, via the spatial regression-based term $b_k^*(\mathbf{u})$, and (ii) the spatial auto- and cross-correlation of the resulting residuals, via the cokriging-based term $e_k^*(\mathbf{u})$.

Since cokriging is an exact interpolator, the cokriging-predicted spatial regression residuals reproduce the original spatial regression residuals at the rain gauges, i.e. $e_k^*(\mathbf{u}_\alpha) = e_k(\mathbf{u}_\alpha),\ \forall k, \alpha$. This entails that the original temporal trend coefficients are also reproduced at the rain gauges, i.e. $b_k^{**}(\mathbf{u}_\alpha) = b_k^*(\mathbf{u}_\alpha) + e_k^*(\mathbf{u}_\alpha) = b_k(\mathbf{u}_\alpha),\ \forall k, \alpha$. In addition, the final estimated temporal trend at any rain gauge $\mathbf{u}_\alpha$ for any time instant t reproduces the original temporal trend value for that instant, i.e. $m_\alpha^*(t) = m_\alpha(t),\ \forall \alpha, t$.

Availability of a set of $(K+1)$ such estimated coefficient values $\{b_k^{**}(\mathbf{u}), \mathbf{u} \in D\}$, $k = 0, \ldots, K$, yields an estimate $\{m_{\mathbf{u}}^*(t),\ \mathbf{u} \in D,\ t \in T\}$ of the temporal trend component over the space time domain, as:

$$m_{\mathbf{u}}^*(t) = \sum_{k=0}^{K} b_k^{**}(\mathbf{u}) f_k(t), \qquad \mathbf{u} \in D,\ t \in T \tag{14}$$

2.3. Stochastic simulation of space–time precipitation

Once the spatiotemporal trend component $\{m_{\mathbf{u}}^*(t),\ t \in T\}$ is estimated at any ungauged location $\mathbf{u}$, stochastic simulation of daily precipitation at any time instant t and at that location $\mathbf{u}$ amounts to simulating realizations of the spatiotemporal residual component $\{R_{\mathbf{u}}(t),\ t \in T\}$ at that location and time instant, and combining them with the previously estimated trend component. In what follows, the collection of all (unknown) time profiles over the study region is denoted as $\{m(\mathbf{u}, t),\ \mathbf{u} \in D,\ t \in T\}$ for the spatiotemporal trend component, and $\{R(\mathbf{u}, t),\ \mathbf{u} \in D,\ t \in T\}$ for the spatiotemporal residual component. This latter component is modeled as a stationary (in space and time) zero mean process. Since any temporal and spatial non-stationarity is accounted for by the spatiotemporal trend component $\{m(\mathbf{u}, t),\ \mathbf{u} \in D,\ t \in T\}$, the assumption of a stationary space–time residual component is more realistic. In the most frequently encountered case of non-Gaussian precipitation data, stationarity applies to Gaussian-transformed residuals (see the description of the simulation method that follows).

Stochastic characterization of the residual component $\{R(\mathbf{u}, t),\ \mathbf{u} \in D,\ t \in T\}$ calls for modeling the spatiotemporal covariance of the (possibly Gaussian-transformed) r-residuals of Eq. (6):

$$\begin{aligned} \mathrm{Cov}\{R(\mathbf{u}, t), R(\mathbf{u}', t')\} &= E\{R(\mathbf{u}, t) R(\mathbf{u}', t')\} \\ &= C_R(\mathbf{u} - \mathbf{u}', t - t') = C_R(\mathbf{h}, \tau) \end{aligned} \tag{15}$$

where $\mathbf{h} = \mathbf{u} - \mathbf{u}'$ denotes a spatial lag vector, and $\tau = t - t'$ denotes a temporal lag. When $\tau = 0$, the covariance $C_R(\mathbf{h}, 0)$ models the spatial correlation of the residuals, which is the same for all time instants. When $\mathbf{h} = 0$, the covariance $C_R(\mathbf{0}, \tau)$ models the temporal correlation of the residuals, which is

the same from one location to another. Scale differences between the space and time domains are accounted for via a geometric anisotropy using a generalized distance metric:

$$\sqrt{(\mathbf{h}/a_1)^2 + (\tau/a_2)^2},$$

where a_1 and a_2 denote (unequal) spatial and temporal correlation length parameters, see Section 3.3 and Kyriakidis and Journel (1999) for details.

In this paper, sequential Gaussian simulation (Deutsch and Journel, 1998) is used to generate a S-member ensemble of spatiotemporal precipitation realizations $\{z^{(s)}(\mathbf{u}, t),\ \mathbf{u} \in D,\ t \in T\},\ s = 1, \ldots, S$; here superscript s denotes the s-th member of the ensemble. Spatiotemporal sequential simulation is similar to simulation from an AR process, whereby a random visiting sequence of locations (in space–time) replaces the natural ordering of time driving the simulation in the case of AR processes.

In sequential simulation, the use of kriging and an invertible quantile transformation procedure guarantee that the original rain gauge data are reproduced at their locations (in space–time) in all simulated realizations. Any missing precipitation values (due to missing observations) at the rain gauges are in-filled via simulation. It should be noted here, that a sole reproduction of the rain gauge data at their locations does not suffice to reproduce the rain precipitation histogram, because the simulated values are many more than the original rain gauge data (since the entire space–time domain is in-filled with such simulated values). The quantile transformation procedure adopted in sequential simulation, however, ensures reproduction (within statistical fluctuations) of the rain gauge precipitation distribution by each simulated realization.

Sequential simulation proceeds in the following steps (for more details, the reader is referred to Deutsch and Journel (1998)):

1. The original rain gauge z-data are transformed to a new set of values that follows a standard Gaussian distribution, using a non-linear, rank-preserving, and invertible quantile transformation procedure. More precisely, any original rain gauge datum $z(\mathbf{u}_\alpha, t_i)$ is transformed to a standard Gaussian deviate $z_G(\mathbf{u}_\alpha, t_i)$ as: $z_G(\mathbf{u}_\alpha, t_i) = G^{-1}[F_Z[z(\mathbf{u}_\alpha, t_i)]]$, where $F_Z[\cdot]$ denotes the cumulative distribution function (cdf) of the rain gauge data, and $G^{-1}[\cdot]$ denotes the inverse standard Gaussian cdf. The reader should not confuse F_Z and G with the temporal design matrix $\mathbf{F}$ of Eq. (4) and the spatial design matrix $\mathbf{G}$ of Eq. (7), respectively. Any estimated trend value $m^*(\mathbf{u}_\alpha, t_i)$ is transformed to a Gaussian deviate $m^*_G(\mathbf{u}_\alpha, t_i)$ according to the cdf of the rain gauge data: $m^*_G(\mathbf{u}_\alpha, t_i) = G^{-1}[F_Z[m^*(\mathbf{u}_\alpha, t_i)]]$. Note that the transformed m^*_G-values do not follow a standard Gaussian distribution; they would, if in the above equation the cdf of the rain gauge data $F_Z[\cdot]$ was replaced by the cdf of the estimated trend values $F_{m^*}[\cdot]$. In particular, the variance of the m^*_G-values is not one, because the trend component quantifies some 'average' spatiotemporal variability of the precipitation process. The Gaussian r_G-residuals at the rain gauges are then obtained as: $r_G(\mathbf{u}_\alpha, t_i) = z_G(\mathbf{u}_\alpha, t_i) - m^*_G(\mathbf{u}_\alpha, t_i)$. It is this set of r_G-data that are assumed stationary, and whose covariance model is specified as $C_{R_G}(\mathbf{h}, \tau)$; in the Gaussian case where no transformation is necessary, $C_{R_G}(\mathbf{h}, \tau) = C_R(\mathbf{h}, \tau)$, with $C_R(\mathbf{h}, \tau)$ given in Eq. (15).
2. A random path is defined for visiting once all simulation grid nodes (in space and time). If one wishes to condition simulated values at time instant t_i only to previously simulated values at time instants before t_i, then the visiting sequence should be such that simulation grid nodes at time t_i are only visited (at random) after any node prior to t_i, and before any node after t_i. This approach was not adopted in this work, because the objective here is simulation in a hindcast mode. Instead, the simulated value at a node at time t_i was conditioned to observed values before and after t_i.
3. At any simulation node $(\mathbf{u}, t_i)$ visited at random along this path:
   (a) Kriging is performed to estimate the mean $r^*_G(\mathbf{u}, t)$ and standard deviation $\sigma^*_{R_G}(\mathbf{u}, t)$ of the local (node-specific) residual Gaussian distribution, conditional to the r_G-data and any previously simulated r_G-values at grid nodes visited before $(\mathbf{u}, t)$. The covariance model $C_{R_G}(\mathbf{h}, \tau)$ is used in this kriging step. At the first node of this random path,

there are no previously simulated r_G-values; the only conditioning information available consists of the rain gauge r_G-data. The corresponding Gaussian transformed trend value $m^*_G(\mathbf{u}, t)$ at this node is added to the mean $r^*_G(\mathbf{u}, t)$ of this distribution, which now becomes the local Gaussian distribution of z_G-values with mean $m^*_G(\mathbf{u}, t) + r^*_G(\mathbf{u}, t)$ and standard deviation $\sigma^*_{R_G}(\mathbf{u}, t)$.

(b) A simulated Gaussian deviate $z^{(s)}_G(\mathbf{u}, t)$ is generated at this node from the above distribution using Monte Carlo drawing, and it is added to the conditioning data set (of rain gauge r_G-data, and previously simulated r_G-values). This is equivalent to first simulating a residual value $r^{(s)}_G(\mathbf{u}, t)$ from the local Gaussian distribution with mean $r^*_G(\mathbf{u}, t)$ and standard deviation $\sigma^*_{R_G}(\mathbf{u}, t)$, and then adding to this simulated value the corresponding estimated trend component $m^*_G(\mathbf{u}, t)$.

4. All simulation grid nodes along the random path are visited sequentially, and the above two steps are repeated at each node.
5. The set of all simulated Gaussian z_G-values is then back-transformed to a set of z-values with a non-Gaussian distribution (that of the original rain gauge z-data). In particular, a simulated precipitation value $z^{(s)}(\mathbf{u}, t)$ is obtained from the corresponding simulated Gaussian deviate $z^{(s)}_G(\mathbf{u}, t)$ as: $z^{(s)}(\mathbf{u}, t) = F_Z^{-1}[G[z^{(s)}_G(\mathbf{u}, t)]]$, where $F_Z^{-1}[\cdot]$ denotes the sample inverse cdf and $G[\cdot]$ denotes the standard Gaussian cdf. The set of backtransformed simulated values constitutes the s-th precipitation realization $\{z^{(s)}(\mathbf{u}, t),\ \mathbf{u} \in D,\ t \in T\}$.

A new realization $\{z^{(s')}(\mathbf{u}, t), \mathbf{u} \in D, t \in T\}$ is generated by repeating all the above steps with a different random path. The set of S alternative simulated realizations $\{z^{(s)}(\mathbf{u}, t), \mathbf{u} \in D, t \in T\}, s = 1, \ldots, S$, provides a model of uncertainty for the unknown precipitation levels in both space and time, which can be used for hydrologic impact assessment studies (Seo et al., 2000; Kyriakidis et al., 2001a).

Because in this work the temporal trend b-coefficients (hence the spatiotemporal trend component) are treated as deterministic, we do not account for the uncertainty in the predicted b-coefficients in the simulation procedure. Such uncertainty stems from the temporal regression of Eq. (4), the spatial regression of Eq. (8), and the cokriging prediction of Eq. (11). Modifications of the sequential simulation procedure described above can be made to account for these extra uncertainty sources, and will be reported in the near future. For a more elaborate procedure, which accounts for the uncertainty in the estimated trend component at ungauged locations via simulation of the spatiotemporal trend component, the reader is referred to Kyriakidis and Journel (2001).

3. Case study

The study domain is a 300×360 km^2 area of the northern California coastal region, which is characterized by complex terrain and significant spatial and seasonal variation in precipitation. Annual precipitation varies from 200 mm/year in the Central Valley (east of the Coastal Range) to over 1300 mm/year in the Santa Cruz Mountains (north of the Monterey Bay). Precipitation in the region generally originates from stratiform clouds due to orographic lifting of the westerly flow over the western slope of the Coastal Range. Occasionally, strong convection embedded within the stratiform clouds generates intense local precipitation.

The rainfall data set used in this study consists of 77 rain gauge precipitation measurements of daily rainfall during the 92 days from Nov 1, 1981 to Jan 31, 1982. The time average of precipitation for these 92 days at each rain gauge is shown in Fig. 1A. The original daily precipitation values constitute a subset of the Cooperative observer (COOP) and first-order precipitation stations, obtained from the National Climate Data Center (NCDC, 1995); for a detailed analysis of California precipitation using this data set, the reader is referred to Pandey et al. (1999). The proportion of rain gauge data above the threshold of 0.25 mm (indicating a wet day) over all 92 days is 0.39. Wet-day precipitation amounts range from 0.25 to 291.38 mm, with a mean of 14.98 mm and a median of 6.35 mm indicating a positively skewed precipitation distribution. The standard deviation and coefficient of variation of the wet-day precipitation amounts is 23.88 and 1.59 mm, respectively, indicating a significant

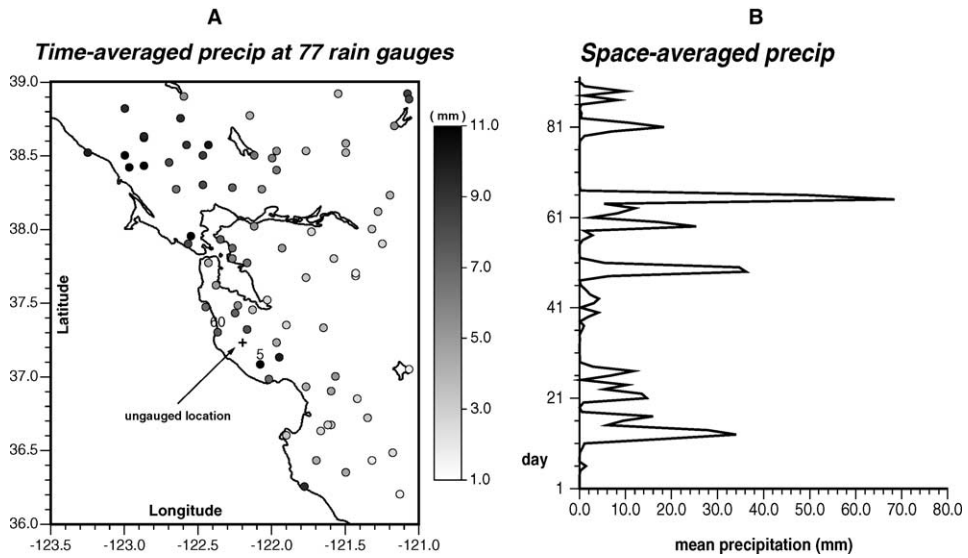

Fig. 1. Time-average of observed daily precipitation at 77 rain gauges during the period from Nov 1, 1981 to Jan 31, 1982 (A), and space-average of precipitation profiles for the same 92 days (B); the cross indicates an ungaged location at which simulated precipitation profiles are shown along with those observed at nearby stations #5 and #60 (see Fig. 10).

spatiotemporal variability. The objective of this study is to generate ensemble predictions of precipitation on a 300×360 grid of cell size 1 km^2 for the period Nov 1, 1981 to Jan 31, 1982, using all relevant information available for this region.

Auxiliary spatial variables used in this study include terrain elevation, and its interaction with specific humidity derived from NCEP/NCAR reanalysis data. A smoothed version of a United States Geological Survey (USGS) digital elevation model was used in this study. The smoothing window of 13×13 km^2 was determined by maximizing the correlation between time-averaged precipitation (Fig. 1A) and collocated smoothed elevation; see Kyriakidis et al. (2001b) for details. Time-averaged specific humidity integrated over $850 - 1000$ hPa was derived by interpolation from the nine NCEP/NCAR reanalysis nodes closest to the study domain, and represents the availability of large-scale water vapor in the lower atmosphere over the period of interest. Areas where the product (interaction in statistical terms) of the above two variables (smoothed elevation and integrated specific humidity) is high, correspond to places where both elevation and moisture availability, hence likelihood of precipitation, is high.

3.1. Parametric local temporal trend models

The first step in the proposed methodology is to establish a set of local temporal trend models of precipitation at each rain gauge, see Section 2.1. To this effect, two elementary profiles are used as temporal precipitation predictors at each rain gauge:

$$\mathbf{f}_0 = [f_0(t_i) = 1,\ i = 1, \ldots, 92]', \text{ and}$$

$$\mathbf{f}_1 = \left[\frac{1}{n} \sum_{\alpha=1}^{n} z(\mathbf{u}_\alpha, t_i),\ i = 1, \ldots, 92 \right]',$$

see Eq. (4). In other words, the spatial average $\mathbf{f}_1$ of the precipitation profiles from the 77 rain gauges (Fig. 1B) is used as the temporal precipitation predictor at each rain gauge. By analogy to simple linear regression, two temporal trend coefficients are available at each rain gauge $\mathbf{u}_\alpha$ (see Fig. 2): an intercept coefficient $b_0(\mathbf{u}_\alpha)$ associated with $\mathbf{f}_0$, and a slope coefficient $b_1(\mathbf{u}_\alpha)$ associated with $\mathbf{f}_1$. Rain gauges with near zero intercept and near unit slope values (see the eastern part of the study domain and the south Bay Area) indicate precipitation profiles very similar to the spatially averaged profile $\mathbf{f}_1$.

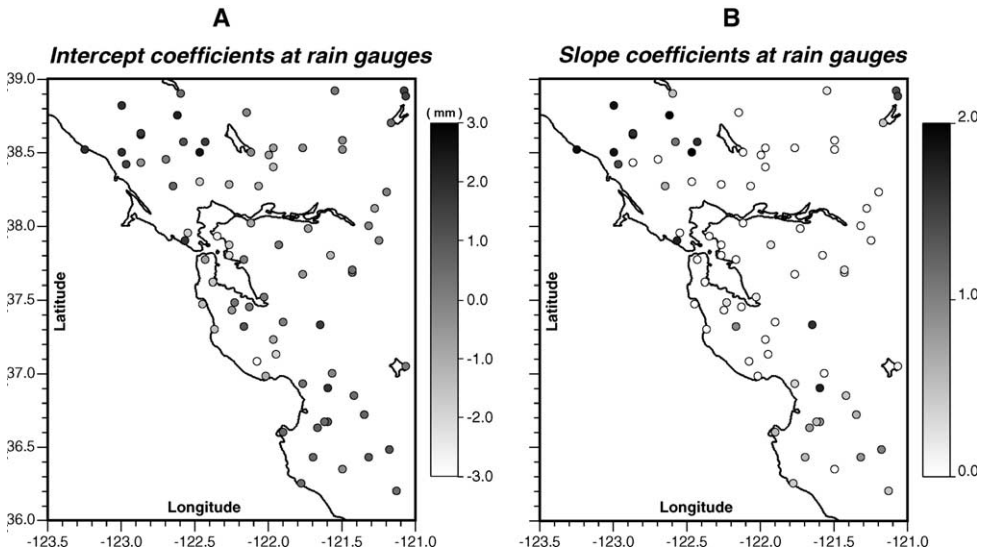

Fig. 2. Coefficients, intercept (A) and slope (B), of local temporal trend models established at the 77 rain gauges.

A measure of how well the spatially averaged precipitation profile $\mathbf{f}_1$ captures local variations at the α-th rain gauge is the coefficient of determination (CD_α) computed as:

$$\mathrm{CD}_\alpha = \frac{\sum_{i=1}^{T_\alpha} [m_\alpha(t_i) - \bar{z}_\alpha]^2}{\sum_{i=1}^{T_\alpha} [z_\alpha(t_i) - \bar{z}_\alpha]^2}$$

where $\bar{z}_\alpha$ denotes the time-averaged precipitation at the α-th rain gauge (shown in Fig. 1A).

The spatial variability of these CD-values is shown in Fig. 3A. The average CD-value is 0.58, with a minimum of 0.09 and a maximum of 0.87 (Fig. 3B), indicating that the proportion of temporal precipitation variance accounted for by the spatially averaged precipitation profile $\mathbf{f}_1$ changes significantly from one rain gauge to another. Precipitation profiles at rain gauges with high CD-values (located in the northern part of the study area and in the Santa Cruz mountains) can be adequately characterized by a linear rescaling of the spatially averaged profile $\mathbf{f}_1$.

3.2. Spatiotemporal trend component

Once a set of local trend models is established at each of the $n = 77$ rain gauge locations, the task is to estimate the spatiotemporal trend component of precipitation $m(\mathbf{u}, t)$, see Eq. (2), at any grid cell $\mathbf{u}$ and for any day t. This task calls for the joint spatial prediction of intercept b_0 and slope b_1 coefficients at any location $\mathbf{u}$ within the study domain D. Joint spatial prediction of intercept b_0- and slope b_1-coefficients is enhanced by accounting for their relation with terrain elevation and its interaction with specific humidity derived from NCEP/NCAR reanalysis data (see Section 2.2).

The rank transform of the window averaged elevation is used as an auxiliary variable in the spatial prediction of intercept b_0-coefficients. Similarly, the rank transform of the product (interaction) of specific humidity with the smoothed terrain elevation is used as an auxiliary variable in the spatial prediction of slope b_1-coefficients. The values of the above ranked transformed predictors at the $n = 77$ rain gauges, together with the (77×1) vector of ones form the three columns of the design matrix $\mathbf{G}_n$ of Eq. (7). The coefficients of determination for the regression of intercept b_0-coefficients (Fig. 2A) on collocated rank-transformed smoothed elevation values and of slope b_1-coefficients (Fig. 2B) on rank-transformed humidity–elevation interaction values are 0.3 and 0.4, respectively, see Eq. (7). Both regression models are statistically significant at the 95% level, but provide marginal benefits for prediction purposes, when

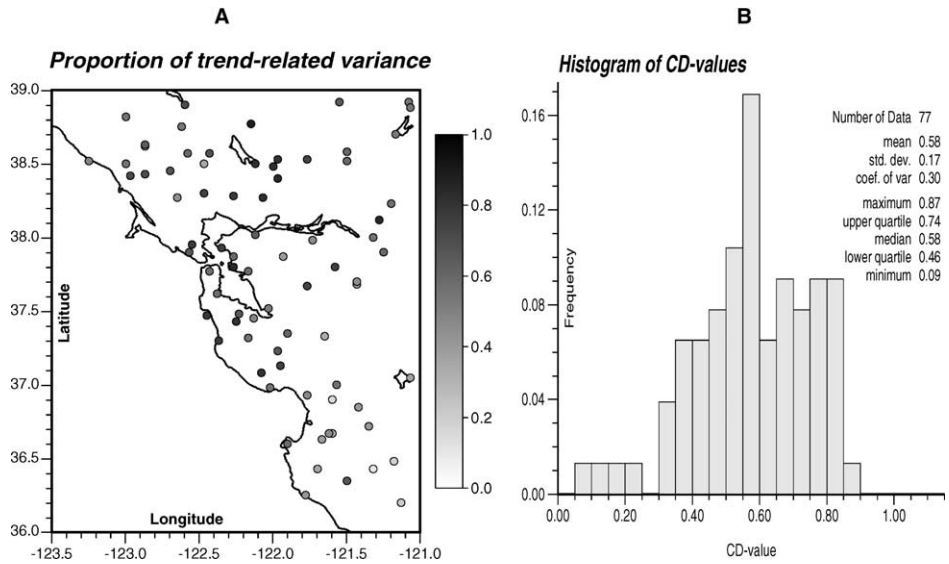

Fig. 3. Proportion of variance, quantified by the coefficient of determination CD_α, of precipitation temporal variability accounted for by local temporal trend models at the 77 rain gauges (A), and histogram of CD_α-values (B).

compared to the impact of the spatial autocorrelation of the residual values themselves; see Goovaerts (1997) for details regarding the effect of auxiliary variables on cokriging predictions.

SCK was used for the joint spatial prediction of the resulting regression residual e_0- and e_1-values, see Section 2.2 and Eqs. (11) and (12). All auto- and cross-variogram functions of these residuals (not shown) were jointly modeled as isotropic using the linear model of coregionalization (LMC), see Wackernagel (1995) for details, and are:

$$\gamma_{E_0}(|\mathbf{h}|) = 0.01[1 - \delta_{|\mathbf{h}|}] + 0.56\left[1 - \exp\left(\frac{-3|\mathbf{h}|^2}{20^2}\right)\right] + 0.80\left[1 - \exp\left(\frac{-3|\mathbf{h}|}{80}\right)\right]$$

$$\gamma_{E_1}(|\mathbf{h}|) = 0.01[1 - \delta_{|\mathbf{h}|}] + 0.07\left[1 - \exp\left(\frac{-3|\mathbf{h}|^2}{20^2}\right)\right] + 0.12\left[1 - \exp\left(\frac{-3|\mathbf{h}|}{80}\right)\right]$$

$$\gamma_{E_0E_1}(|\mathbf{h}|) = 0.00[1 - \delta_{|\mathbf{h}|}] + 0.19\left[1 - \exp\left(\frac{-3|\mathbf{h}|^2}{20^2}\right)\right] + 0.09\left[1 - \exp\left(\frac{-3|\mathbf{h}|}{80}\right)\right]$$

where $|\mathbf{h}|$ denotes the norm of vector $\mathbf{h}$, $\gamma_{E_0}(|\mathbf{h}|)$ denotes the semivariogram model for the e_0 residuals of intercept b_0-coefficients from the regression on the rank transform of the window averaged elevation, $\gamma_{E_1}(|\mathbf{h}|)$ denotes the semivariogram model for the e_1 residuals of slope b_1-coefficients from the regression on the rank transform of the product (interaction) of specific humidity with smoothed terrain elevation, and $\gamma_{E_0E_1}(|\mathbf{h}|)$ denotes the cross-semivariogram between these two sets of residual values. All models are comprised from three nested structures: a nugget effect $1 - \delta_{|\mathbf{h}|}$ (where $\delta_{|\mathbf{h}|} = 1$ if $|\mathbf{h}| = 0$, zero otherwise), a Gaussian structure with effective range (distance at which 95% of total sill is reached) 20 km, and an exponential structure with effective range 80 km.

Note that the correlation coefficient $\rho_{E_0E_1}(0)$ between the two residual data sets can be deduced from the sill $\gamma_{E_0E_1}(\infty)$ of the cross-semivariogram

model as:

$$\rho_{E_0E_1}(0) = \gamma_{E_0E_1}(\infty)/\sqrt{\gamma_{E_0}(\infty)\gamma_{E_1}(\infty)}$$

where $\gamma_{E_0}(\infty)$ denotes the semivariogram sill (variance) of the e_0-residuals and $\gamma_{E_1}(\infty)$ denotes the semivariogram sill (variance) of the e_1-residuals. In this case,

$$\rho_{E_1E_2}(0) = -(0.19 + 0.09)/\sqrt{(0.01 + 0.56 + 0.8)(0.01 + 0.07 + 0.12)} = -0.53,$$

which is equal to the sample correlation coefficient between the two sets of e-residual values.

The maps of estimated temporal trend coefficients, intercept b_0^{**}-values and slope b_1^{**}-values resulting from Eq. (13) are shown in Fig. 4A and B. Note that (co)kriging is an exact interpolator, which implies that regression residual e_k-values, hence temporal trend coefficient b_k-values, are reproduced at their respective rain gauge locations. Also note the negative correlation between the estimated coefficients: high-valued b_0 intercept areas (dark-colored pixels in Fig. 4A) generally correspond to low-valued b_1 slope areas (light-colored pixels in Fig. 4B).

Using the above estimated intercept $b_0^{**}(\mathbf{u})$ and slope $b_1^{**}(\mathbf{u})$ coefficients at any grid cell $\mathbf{u}$, one can estimate the spatiotemporal trend component $m^*(\mathbf{u}, t)$ at any grid cell $\mathbf{u}$ and any day t using Eq. (14). Maps of this estimated temporal trend component for Nov 12 and 13, 1981 are shown in Fig. 5A and B. Note the higher trend values for Nov 13 as compared to those of Nov 12, especially over the Santa Cruz mountains and over the south tip of the Coastal Range.

3.3. Stochastic simulation of space–time precipitation

Stochastic simulation of daily precipitation in space and time amounts to combining to the estimated spatiotemporal trend component $\{m^*(\mathbf{u}, t),\ \mathbf{u} \in D,\ t \in T\}$ with a realization of the spatiotemporal residual component $\{R(\mathbf{u}, t),\ \mathbf{u} \in D,\ t \in T\}$ of Eq. (2). Simulation of the spatiotemporal residual component is performed using sequential Gaussian simulation (see Section 2.3 for details). Stochastic simulation in space and time calls for a spatiotemporal semivariogram model of the Gaussian-transformed r_G-residuals. In this work, a single transformation common to all spatial locations and all time instants is adopted.

The (standardized to unit sill) space–time semivariogram model $\gamma_{R_G}(|\mathbf{h}|, \tau)$ (not shown) adopted for these Gaussian-transformed spatiotemporal

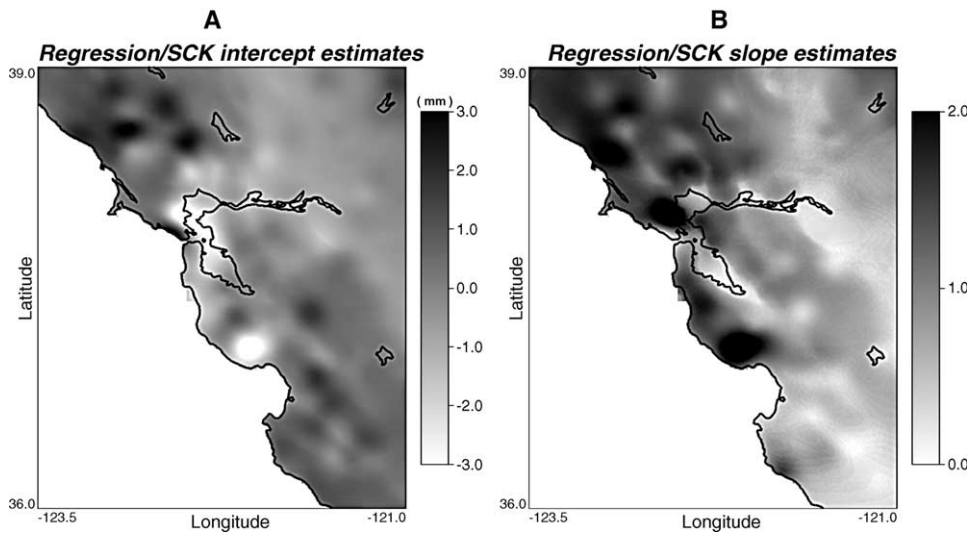

Fig. 4. Maps of estimated temporal trend coefficients, intercept (A) and slope (B), derived respectively by regression on elevation and its interaction with NCEP/NCAR specific humidity, followed by simple cokriging (SCK) of the resulting residuals.

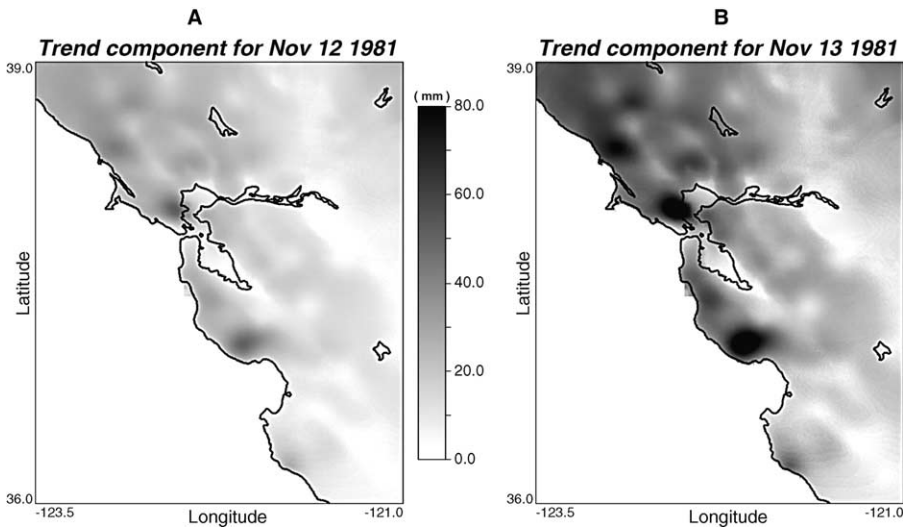

Fig. 5. Maps of precipitation spatiotemporal trend component for Nov 12 (A) and Nov 13 (B) 1981.

r_G-residuals is:

$$\gamma_{R_G}(|\mathbf{h}|,\tau) = 0.30[1-\delta_{|\mathbf{h}|,\tau}] + 0.35\left[1-\exp\left(-3\sqrt{\left(\frac{|\mathbf{h}|}{95}\right)^2+\left(\frac{\tau}{\varepsilon}\right)^2}\right)\right] + 0.35\left[1-\exp\left(-3\sqrt{\left(\frac{|\mathbf{h}|}{95}\right)^2+\left(\frac{\tau}{2}\right)^2}\right)\right] \quad (16)$$

where $\delta_{|\mathbf{h}|,\tau}$ is defined as: $\delta_{|\mathbf{h}|,\tau} = 1$, if $|\mathbf{h}| = \tau = 0$, zero otherwise, and ε denotes a very small number that is used to force the second nested structure to appear as a pure nugget effect in time.

For this particular set of Gaussian transformed r_G-residuals, temporal correlation is very weak, since only 35% of the residual temporal variability is non-random with a correlation period of two days (third nested semivariogram component). Spatial variability exhibits a correlation length of 95 km with a 30% purely random variability (first nested semivariogram structure). This implies that once the spatiotemporal trend component is removed, the residual component (which explains, on average, 42% of spatiotemporal variability) exhibits significant spatial but weak temporal autocorrelation.

Using the space–time semivariogram model of Eq. (16) of the Gaussian-transformed r_G-residual values, sequential simulation (see Section 2.3) was used to generate a set of $S = 30$ alternative realizations of daily precipitation over the 300×360 grid of cell size 1 km^2 for the 92 days from Nov 1, 1981 to Jan 31, 1982. Two of these realizations for Nov 12, 1981 and two for Nov 13, 1981 are shown in Figs. 6A and B and 7A and B. Conditioning entails that areas around high (low) rain gauge precipitation values (see Figs. 6C and 7C) appear also as areas of high (low) precipitation in all simulated realizations.

It should be noted that the proposed simulation approach does not reproduce the histogram of precipitation measurements at any particular sub-region or time instant. Instead, it reproduces the histogram of all precipitation measurements (not that of rainfall depth) over the entire study region and time period of interest. In other words, the Gaussian transformation and back-transformation employed in this case study is global, i.e. considers all data in space and time. Reproduction of time- or region-specific precipitation histograms can be imposed by progressively applying the quantile transformation procedure described in Section 2.3 at simulation grid nodes that lie further away from the informed locations and time instants. Such a progressive transformation does not ruin the data-exactitude property of conditional

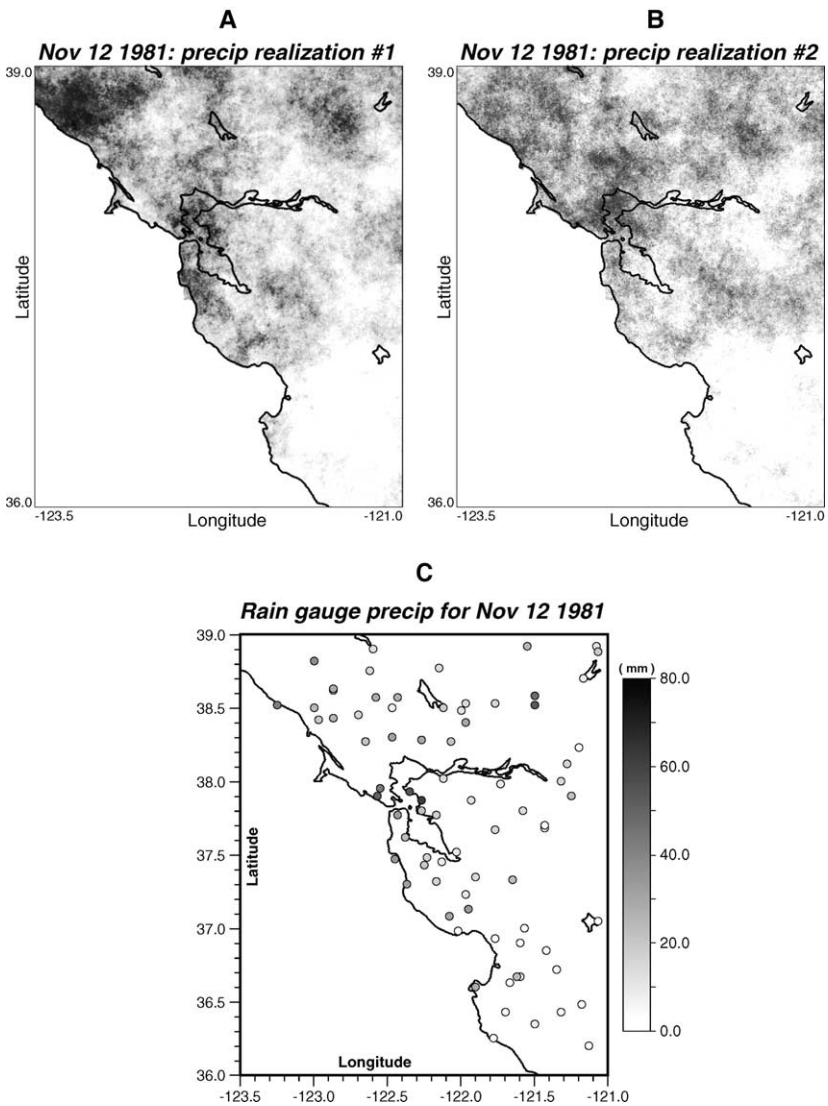

Fig. 6. Two (out of 30) synthetic precipitation fields for Nov 12, 1981 (A and B) generated by conditional stochastic simulation, along with the contemporaneous rain gauge data (C).

simulations (Journel and Xu, 1994; Deutsch and Journel, 1998), and thus was adopted in this case study. All subsequent figures pertain to such transformed realizations.

The reproduction of the histograms of sample precipitation recorded at the 77 rain gauges at Nov 12 and 13, 1981, by the histograms of five precipitation realizations for each of these days is shown via the quantile–quantile plots of Fig. 8A and B. A plot aligned along the first bisector implies two nearly identical distributions. The semivariogram reproduction for Nov 12 and 13, 1981 is shown in Fig. 8C and D; the sample precipitation semivariograms are well approximated by the semivariograms of the five precipitation realizations. Conditionally simulated daily precipitation realizations thus provide realistic synthetic representations of the true (unknown) precipitation field, insofar as they reproduce the histogram and semivariogram of observed rain gauge data.

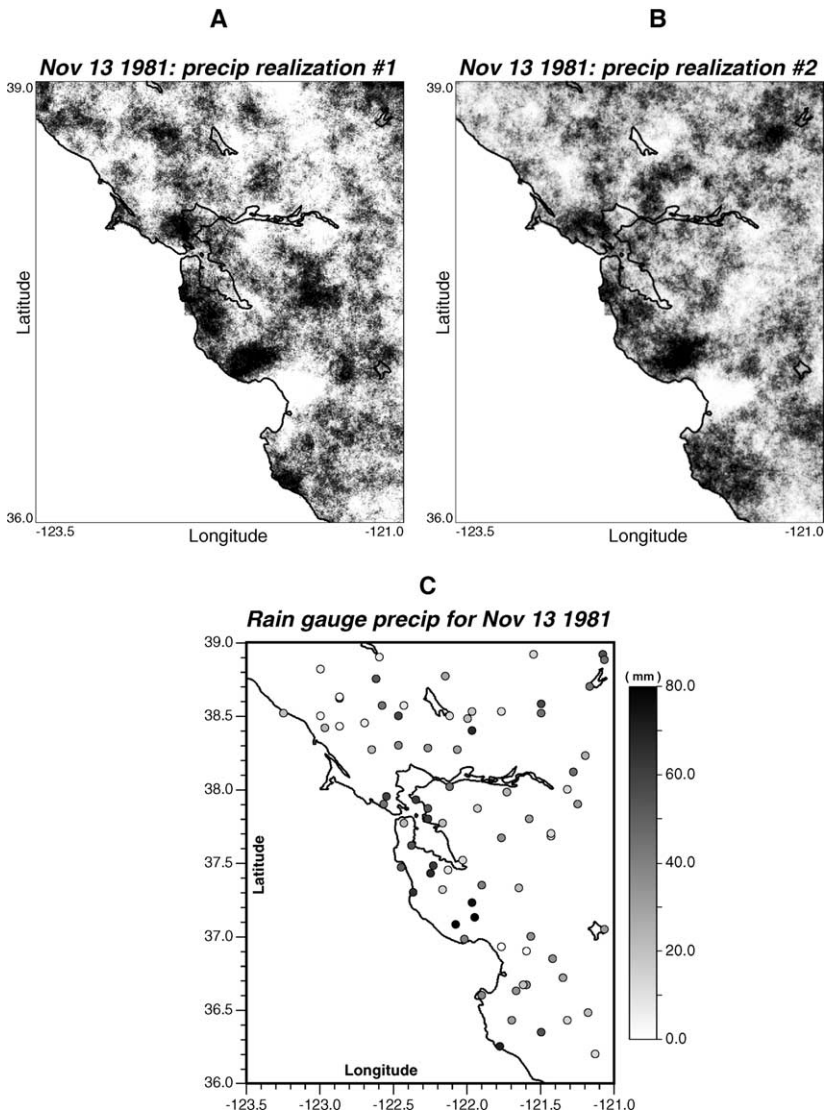

Fig. 7. Two (out of 30) synthetic precipitation fields for Nov 13, 1981 (A and B) generated by conditional stochastic simulation, along with the contemporaneous rain gauge data (C).

A summary of simulated precipitation at each grid cell is provided by the (ensemble) average and standard deviation of the simulated values at that cell. Maps of ensemble averages of simulated precipitation for Nov 12 and Nov 13, 1981 are given in Fig. 9A and B. Note the high precipitation amounts in the Santa Cruz mountains for Nov 13, a pattern consistent with that deduced from the contemporaneous rain gauge data (Fig. 7C). Maps of ensemble standard deviations of simulated precipitation for Nov 12 and Nov 13, 1981 are given in Fig. 9C and D. Note the increased standard deviation values for Nov 13 with respect to those observed for Nov 12, as well as the small standard deviation values near rain gauges, which indicate less spatial uncertainty around these locations. Other summary maps, such as maps of probability that precipitation exceeds a critical threshold used, say, in flood warnings, can be also generated from the ensemble of synthetic precipitation fields. It should be noted here that the depicted

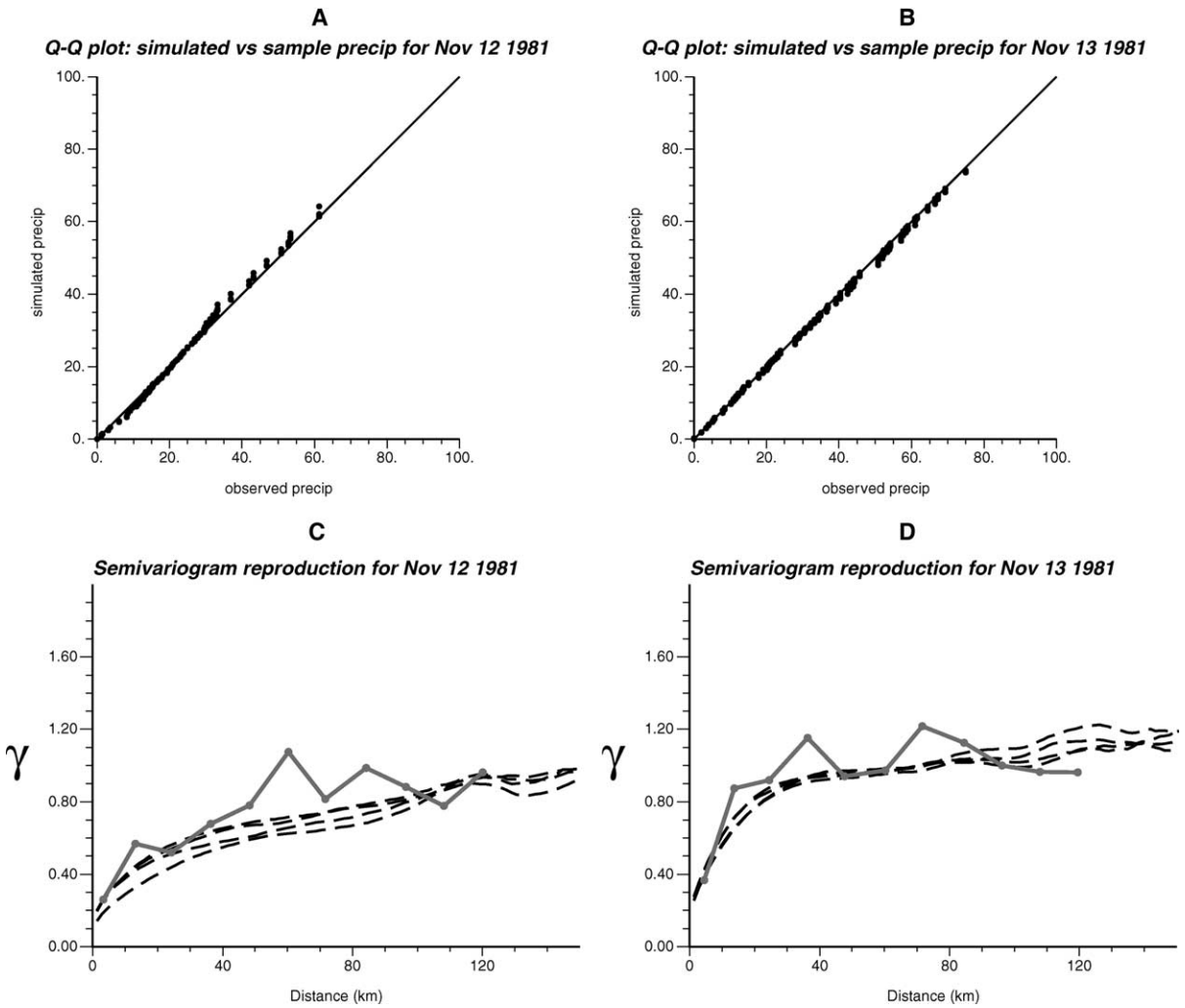

Fig. 8. Reproduction of observed precipitation histogram (A) and semivariogram (B) from five precipitation realizations for Nov 12 and 13, 1981 (solid line: semivariogram of observed precipitation; dashed lines: semivariograms of simulated precipitation realizations).

standard deviation maps account only for the uncertainty in the cokriging predictions of the residual component (see Eq. (11)), and thus should be regarded as conservative measures of spatial uncertainty (see related discussion at the end of Section 2.3).

Last, we compare the simulated precipitation profiles at the ungauged location shown (with a cross) in Fig. 1A, with precipitation profiles at two nearby rain gauges #5 and #60, all located in the same mountainous region. The choice of the ungauged location was based on: (i) its proximity to rain gauges with significant temporal variability in their recorded precipitation profiles, and (ii) the fact that the Santa Cruz mountains have a strong orographic influence on local precipitation. Alternatively, we could have used for comparison the precipitation profile of a rain gauge that was not included in the sample set during the analysis. Since our goal is not to validate the method, but simply to illustrate a set of simulated temporal precipitation profiles, we did not opt for this alternative.

The set of 30 simulated profiles, and their ensemble average, at the ungauged location is shown in Fig. 10B. The precipitation profiles at the two nearby

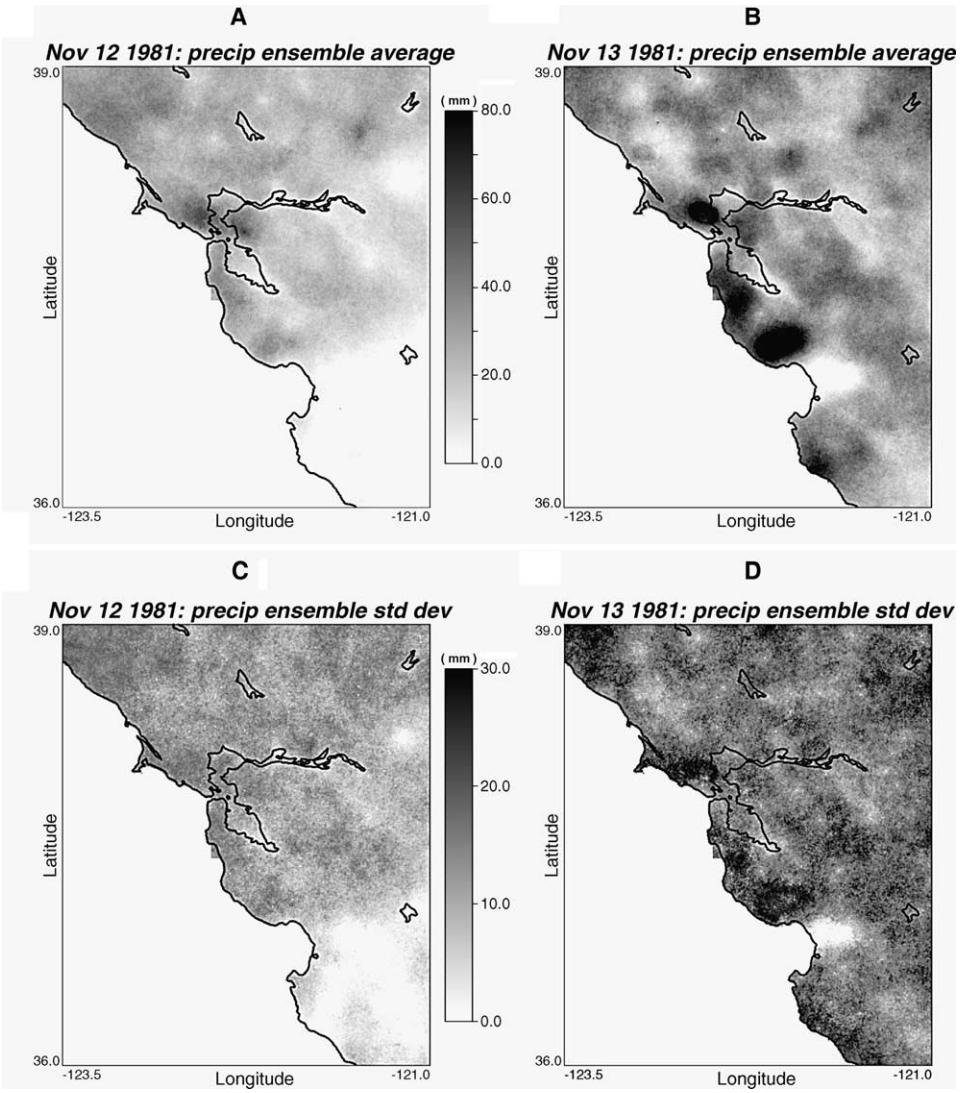

Fig. 9. Precipitation ensemble average (A and B) and ensemble standard deviation (C and D) for Nov 12 and 13, 1981, computed from 30 synthetic precipitation fields generated via conditional stochastic simulation.

rain gauges are shown in Fig. 10A–C. One can appreciate the similarity of the simulated precipitation profiles to the two rain gauge profiles. Note the common rainfall intermittence pattern exhibited by all profiles, and the similarity of the ensemble precipitation average profile (solid line of Fig. 10B) to those of the nearby rain gauges. The average correlation coefficient between the simulated precipitation profiles and the precipitation profile of rain gauge #5 is 0.73 with a standard deviation of 0.16. Similarly, that average correlation coefficient with rain gauge #60 is 0.72 with a standard deviation of 0.16. The ensemble average precipitation profile has correlation coefficient 0.91 with the precipitation profile at rain gauge #5, and 0.89 with that at rain gauge #60. This latter comparison of temporal profiles of simulated and observed precipitation corroborates the fact that daily precipitation realizations generated via the proposed methodology constitute a realistic synthetic representation of the true (unknown) precipitation field.

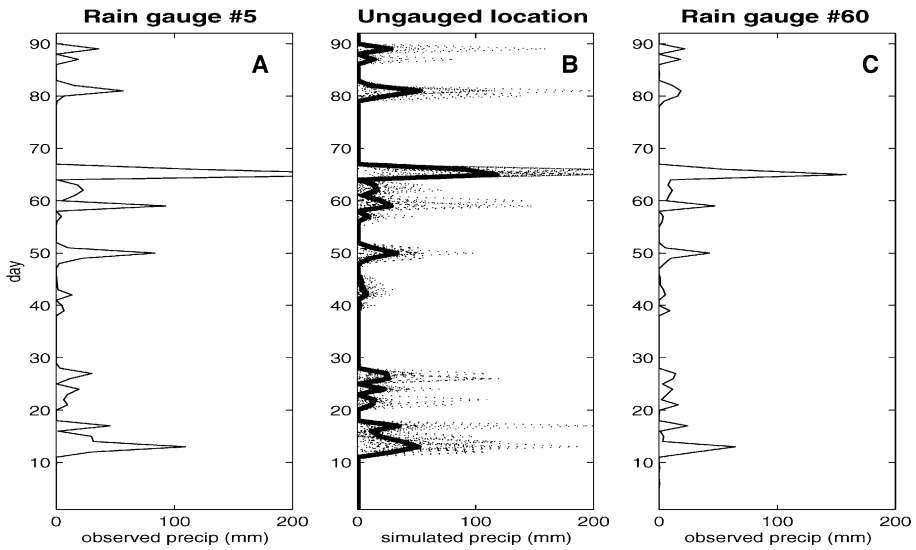

Fig. 10. Illustration of the variability of simulated precipitation at the ungauged location shown in Fig. 1B: 30-member ensemble of simulated daily precipitation profiles at this location (dotted lines) and their ensemble average (thick solid line), A–C: observed precipitation profiles at nearby rain gauges.

4. Summary and discussion

A framework for stochastic spatiotemporal simulation of daily precipitation in a hindcast mode has been presented in this paper. Observed daily precipitation levels are viewed as a joint realization of a collection of spatially correlated time series, thus capitalizing on the typically better informed time domain. The spatiotemporal daily precipitation field is decomposed into a deterministic trend and a stochastic residual component. Parametric temporal trend models are established at all rain gauges, independently from one rain gauge to another, and their parameters are (co)regionalized in space to yield an estimate of the space–time trend component at any location for any day. The joint spatial prediction of such temporal trend coefficients accounts for their covariation in space, as well as for their relation with ancillary spatial information. Simulated realizations of daily precipitation in space and time are obtained by generating alternative realizations of the spatiotemporal residual component (modeled as stationary in space and time), and combining them with the estimated trend component.

The case study illustrated the applicability of the proposed approach using daily precipitation measurements from 77 rain gauges in northern California during 92 days (no missing values) from Nov 1, 1981 to Jan 31, 1982. The spatial average of the 77 precipitation profiles was used as the precipitation predictor, in order to evaluate local departures from this average profile at each rain gauge. The spatial variability of the parameters (intercept and slope) of 77 local (rain gauge specific) temporal linear regression models indicated that there are significant departures of individual rain gauge precipitation profiles from the average profile over the study region. The proportion of temporal precipitation variance at each rain gauge, quantified by the coefficient of determination, ranged from a minimum of 0.09 to a maximum of 0.87 with an average of 0.58; this indicates that there are rain gauges whose precipitation profiles differ dramatically from the spatially averaged profile.

The temporal regression parameters were regionalized (interpolated) on a 300×360 grid of cell size $1\,\mathrm{km}^2$ over the study region. Their spatial interpolation accounted for the effect of ancillary information derived from terrain elevation and NCEP/NCAR reanalysis data of lower-atmosphere moisture availability in the region. The contribution of ancillary information was incorporated via spatial regression models (yet with marginal effect, since the respective coefficients of determination were

0.3 and 0.4) established from collocated intercept b_0 and (rank transformed) elevation values, as well as from collocated slope b_1 and (rank transformed) specific humidity values. The residuals from these spatial regression models were auto- and cross-correlated in space with correlation lengths up to 80 km. Cokriging of these residual values on the same grid as above, yielded interpolated temporal trend model parameters, and consequently estimates of the trend component at each grid node. The rain gauge precipitation residuals (from the corresponding temporal trend models) showed significant spatial but weak temporal correlation. The correlation of these spatiotemporal residuals was modeled as stationary (in both space and time), with respective correlation lengths 95 km and 2 days.

Stochastic conditional simulation was adopted for generating 30 synthetic realizations of daily precipitation on a 300×360 grid of cell size 1 km^2 over the study region. Simulated precipitation realizations reproduced the histogram and semivariogram model of the rain gauge data. In addition, simulated precipitation profiles compared well with observed profiles at nearby rain gauges. The ensemble average and standard deviation fields were also provided as a summary of the 30 synthetic precipitation realizations.

It should be noted, however, that ensemble average fields do not reproduce the statistical properties (histogram, semivariogram) of the rain gauge data; they do reproduce rain gauge precipitation data at their locations, but provide a smooth picture of the spatial distribution of daily precipitation. This latter smoothing characteristic, which also entails misrepresentation of the sample proportion of extreme precipitation events, could be detrimental in hydrologic modeling. Similarly, ensemble standard deviation fields do not provide a measure of joint spatial uncertainty regarding the unknown precipitation value at two or more locations *simultaneously*. Consequently, such fields cannot be used for deriving a measure of uncertainty regarding predictions of hydrological models, e.g. rainfall-runoff models, due to uncertain input forcing; this latter goal is achieved via Monte Carlo simulation. For a detailed discussion regarding the problems associated with ensemble average and standard deviation fields, the reader is referred to Deutsch and Journel (1998).

The proposed approach could be expanded to account for longer periods of dry days, by first simulating a space–time realization of rainfall occurrence and then simulating a space–time realization of rainfall amounts. Realizations of the rainfall amounts process would be generated only in those grid cells at which rainfall was simulated as occurring (wet cells). Results from this latter extension, which enables modeling both patterns of precipitation occurrence and amounts in space and time, will be reported in the near future.

Another research avenue currently being explored pertains to the incorporation of regional climate model predictions of precipitation in the mapping process. Preliminary results show that such regional model predictions contribute significantly to the elucidation of spatiotemporal patterns of precipitation that cannot be discerned from the sparse network of rain gauge measurements in the region. Scale differences between such model predictions and rain gauge measurements can be addressed using a variant of block (co)kriging (Kyriakidis, 2004).

Synthetic conditional simulations of precipitation generated via the proposed approach provide an initial (yet faithful to the rain gauge measurements and their statistics) model of uncertainty regarding unknown daily precipitation levels in both space and time. Such an uncertainty model can be used in a risk analysis context to study the effect of uncertain precipitation forcing on hydrologic impact assessment investigations.

Acknowledgements

The authors would like to thank Dr Demetris Koutsoyiannis and an anonymous reviewer, whose comments greatly improved the original manuscript. This work was partially funded by NASA grant NS-2791. Work at the Department of Energy is under contract DE-AC03-76SF00098.

References

Bierkens, M.F., Knotters, M., Hoogland, T., 2001. Space–time modeling of water table depth using a regionalized time series model and the Kalman filter. Water Resources Research 37, 1277–1290.

Bras, R.L., Rodríguez-Iturbe, I., 1985. Random Functions and Hydrology, Addison Wesley, Reading, MA.

Chandler, R.E., Wheater, H.S., 2002. Analysis of rainfall variability using generalized linear models: a case study from the west of Ireland. Water Resources Research 38 10-1-10-11.

Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide, second ed., Oxford University Press, New York.

Entekhabi, D., Asrar, G.R., Betts, A.K., Beven, K.J., Bras, R.L., Duffy, C.J., Dunne, T., Koster, R.D., Lettenmaier, D.P., McLaughlin, D.B., Shuttleworth, W.J., van Genuchten, M.T., Wei, M.-Y., Wood, E.F., 1999. An agenda for land surface hydrology research and a call for the second international hydrological decade. Bulletin of the American Meteorological Society 80, 2043–2058.

Giorgi, F., Mearns, L.O., 1991. Approaches to the simulation of regional climate change: a review. Reviews of Geophysics 29, 191–216.

Goovaerts, P., 1993. Spatial orthogonality of the principal components computed from coregionalized variables. Mathematical Geology 23(5), 281–302.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation, Oxford University Press, New York.

Hisdal, H., Tveito, O.E., 1992. Generation of runoff series at ungauged locations using empirical orthogonal functions in combination with kriging. Stochastic Hydrology and Hydraulics 6, 255–269.

Hutchinson, M.F., 1995. Stochastic weather models from ground-based data. Agricultural and Forest Meteorology 73, 237–264.

Johnson, G.L., Daly, C., Taylor, G.H., Hanson, C.L., 2000. Spatial variability and interpolation of stochastic weather simulation model parameters. Journal of Applied Meteorology 39, 778–796.

Journel, A.G., Xu, W., 1994. Posterior identification of histograms conditional to local data. Mathematical Geology 22, 323–359.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77, 437–471.

Kim, J., Soong, S.-T., 1996. Simulation of a precipitation event in the western United States. In: Ghan, S.J., Pennel, W.T., Peterson, K.L., Rykiel, E., Scott, M.J., Vail, L.W. (Eds.), Regional Impacts of Global Climate Change, Battelle Press, pp. 73–84.

Kim, J., Miller, N.L., Guetter, A.K., Georgakakos, K.P., 1998. River flow response to precipitation and snow budget in California during the 1994/95 winter. Journal of Climate 11, 2376–2386.

Kyriakidis, P.C., 2004. A geostatistical framework for area-to-point spatial interpolation. Geographical Analysis 36, 1–3.

Kyriakidis, P.C., Journel, A.G., 1999. Geostatistical space–time models: a review. Mathematical Geology 31, 651–684.

Kyriakidis, P.C., Journel, A.G., 2001. Stochastic modeling of atmospheric pollution: a spatial time series approach. Part I: theory. Atmospheric Environment 35, 2331–2337.

Kyriakidis, P.C., Miller, N.L., Kim, J., 2001a. Uncertainty propagation of regional climate model precipitation forecasts to hydrologic impact assessment. Journal of Hydrometeorology 2, 140–160.

Kyriakidis, P.C., Kim, J., Miller, N.L., 2001b. Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics. Journal of Applied Meteorology 40, 1855–1877.

Miller, N.L., Kim, J., 1996. Numerical prediction of precipitation and river flow over the Russian River watershed during the January 1995 California storms. Bulletin of the American Meteorological Society 77, 101–105.

National Climate Data Center (NCDC), 1995. Cooperative Summary of the Day, CDROM set, National Climate Data Center, Federal Bldg., 151 Patton Ave., Ashville, North Carolina.

Pandey, G.R., Cayan, D.R., Georgakakos, K.P., 1999. Precipitation structure in the Sierra Nevada of California during winter. Journal of Geophysical Research 104(D10), 12019–12030.

Rao, A.R., Hsieh, C.H., 1991. Estimation of variables at ungauged locations by empirical orthogonal functions. Journal of Hydrology 123, 51–67.

Sauquet, E., Krasovskaia, I., Leblois, E., 2000. Mapping mean monthly runoff pattern using EOF analysis. Hydrology and Earth System Sciences 4(1), 79–93.

Searle, S.R., 1971. Linear Models, Wiley, New York.

Seo, D.-J., Perica, S., Welles, E., Shaake, J.C., 2000. Simulation of precipitation fields from probabilistic quantitative precipitation forecast. Journal of Hydrology 239, 203–229.

Thiébaux, H.J., 1997. The power of duality in spatial-temporal estimation. Journal of Climate 10, 567–573.

von Storch, H., Zwiers, F.W., 1999. Statistical Analysis in Climate Research, Cambridge University Press, Cambridge.

Wackernagel, H., 1995. Multivariate Geostatistics, Springer, Berlin.

Wilks, D.S., 1998. Multisite generalization of a daily stochastic precipitation generation model. Journal of Hydrology 210, 178–191.